Replica Symmetry Breaking in Compressive Sensing

Ali Bereyhi*, Ralf Müller*, Hermann Schulz-Baldes†
*Institute for Digital Communications (IDC), †Department of Mathematics,
Friedrich Alexander University (FAU), Erlangen, Germany
Email: ali.bereyhi@fau.de, ralf.r.mueller@fau.de, schuba@mi.uni-erlangen.de

***Abstract*—For noisy compressive sensing systems, the asymptotic distortion with respect to an arbitrary distortion function is determined when a general class of least-square based reconstruction schemes is employed. The sampling matrix is considered to belong to a large ensemble of random matrices including i.i.d. and projector matrices, and the source vector is assumed to be i.i.d. with a desired distribution. We take a statistical mechanical approach by representing the asymptotic distortion as a macroscopic parameter of a spin glass and employing the replica method for the large-system analysis. In contrast to earlier studies, we evaluate the general replica ansatz which includes the RS ansatz as well as RSB. The generality of the solution enables us to study the impact of symmetry breaking. Our numerical investigations depict that for the reconstruction scheme with the "zero-norm" penalty function, the RS fails to predict the asymptotic distortion for relatively large compression rates; however, the one-step RSB ansatz gives a valid prediction of the performance within a larger regime of compression rates.**

I. Introduction

The vector-valued linear system

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{z} \tag{1}$$

describes a sampling system in which the source vector $\boldsymbol{x}_{n\times 1} \in \mathbb{X}^n$ with $\mathbb{X} \subseteq \mathbb{R}$ is linearly measured by the sampling matrix $\mathbf{A}_{k\times n} \in \mathbb{R}^{k\times n}$ and corrupted by zero-mean additive white Gaussian noise $\boldsymbol{z}_{k\times 1} \sim \mathcal{N}(\mathbf{0}, \lambda_0 \mathbf{I})$. The source vector is reconstructed from the observation vector $\boldsymbol{y}_{k\times 1}$ using the least-square based reconstruction scheme with

$$\mathbf{g}(\boldsymbol{y}) \coloneqq \arg\min_{\boldsymbol{v}\in\mathbb{X}^n} \left[\frac{1}{2\lambda}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{v}\|^2 + u(\boldsymbol{v})\right] \tag{2}$$

for some general penalty function $u(\cdot)$ and tuning factor λ. The reconstruction scheme in (2) can be considered as a Maximum-A-Posteriori (MAP) estimator which postulates the prior distribution to be proportional to $e^{-u(\boldsymbol{x})}$ and the noise variance to be λ. The optimality of the scheme, therefore, depends on the choice of $u(\cdot)$ and λ. In compressive sensing, the source vector is sparse meaning that it contains a certain number of zero entries [1], [2]. The typical choice for the penalty function in this case is an ℓ_{p}-norm. Different choices of p result in various levels of optimality and complexity which always contain a tradeoff in between; the better the scheme performs, the more complex it is. For noisy sampling systems, the performance of the reconstruction scheme is quantified by the average distortion which reads

$$\mathsf{D}_n = \frac{1}{n}\sum_{j=1}^{n} \mathsf{d}(x_j; \hat{x}_j). \tag{3}$$

for some general distortion function $\mathsf{d}(\cdot;\cdot) : \mathbb{X}\times\mathbb{X} \mapsto \mathbb{R}$, and $\hat{\boldsymbol{x}}_{n\times 1} = \mathbf{g}(\boldsymbol{y})$. In the literature, the most trivial choices for ℓ_{p} are the ℓ_2-norm, ℓ_1-norm and zero-norm which respectively correspond to the "linear", "LASSO" [3] and "zero-norm" reconstruction schemes. The former two choices of p result in convex optimization problems which make them computationally feasible. The latter scheme, however, confronts a non-convex and computationally unfeasible problem. We are interested in studying the asymptotic performance of the general reconstruction scheme given in (2) when the dimensions grow large. The analysis strategy in this case is to consider a random sampling matrix and determine the average distortion for a given realization of it. In this case, the asymptotic performance is evaluated by taking the expectation over the matrix distribution first, and then, taking the limit $n, k \uparrow \infty$. This task is not trivial for most cases of the function $u(\cdot)$ and the support $\mathbb{X}$, and therefore, many analytical methods fail. An alternative approach is based on statistical mechanics in which the asymptotics of the sampling system are represented as macroscopic parameters of a spin glass [4]. In this paper, we take the latter approach and invoke the replica method to study the asymptotics of the reconstruction scheme given in (2).

Replica Method and its Applications

The replica method is a nonrigorous but effective method developed in the physics literature to study spin glasses. Although the method lacks rigorous mathematical proof in some particular parts, it has been widely accepted as an analysis tool and utilized to investigate a variety of problems in applied mathematics, information processing, and coding [5]–[10]. Regarding multiuser communication systems, the method was initially employed by Tanaka to investigate the asymptotic performance of randomly spread CDMA multiuser detectors [11]. For communication systems of form (1) with an independent and identically distributed (i.i.d.) matrix, the authors in [12] considered a class of postulated minimum Mean Square Error (MSE) estimators and extended the earlier analyses to a larger set of input distributions. The study, moreover, justified the decoupling property of the postulated

This work was supported by the German Research Foundation, Deutsche Forschungsgemeinschaft (DFG), under Grant No. MU 3735/2-1.

minimum MSE which was earlier conjectured in [13] and indicates that the pair of input-output symbols are asymptotically converging in distribution to the input-output symbols of an equivalent single-user system. The characteristics of the equivalent system were then determined through the replica analysis. Due to the similarity between the MAP estimation and sampling systems' reconstruction schemes, the replica method has been further used to study compressive sensing [14], [15]. The authors of [16] extended the scope of the decoupling property to a large class of MAP estimators by representing the MAP estimator as the limit of a sequence of minimum MSE etimators and using the replica results of [12]. The result was then employed to study the asymptotics of ℓ_2-, ℓ_1- and zero-nerm based reconstructions in compressive sensing systems. The asymptotic MSE of regularized least-square reconstruction schemes was, moreover, determined in [17] for a wider range of matrices. In [18], the problem of support recovery was considered where the authors determined the asymptotic input-output information rate and support recovery error for a class of sampling systems. The aforementioned studies were considered under the Replica Symmetry (RS) assumption which assumes the equivalent spin glass to have some symmetric properties. Although the RS assumption has been successful in tracking some solutions, there exist several examples in which it clearly fails. In [19], the authors showed that the earlier RS-based investigations of vector precoding in [20] clearly violates the theoretically rigorous lower bound for some example of lattice precoding. They, therefore, employed Parisi's scheme of Replica Symmetry Breaking (RSB) [21] to determine a more general ansatz through the replica analysis. The result depicted that the performance prediction via one-step of RSB is consistent with the theoretical bounds given in the literature. Inspired by [19], the MAP estimator was investigated in [22] under RSB and it was shown that the RS decoupling property reported in [16] holds in a more general form under the RSB assumption. The investigations of the least-square error precoding also has shown several examples in which the RS assumption results in a theoretically invalid solution, and therefore, the RSB ansätze were needed for assessing the performance [23]. Regarding the compressive sensing systems, the stability analysis of ℓ_{p}-norm based reconstruction schemes in [24] for the noiseless sampling systems has shown that in contrast to the convex cases of ℓ_2- and ℓ_1-norm, the RS ansatz for the zero-norm based scheme is not locally stable against perturbations that break the symmetry of the replica correlation matrix. The fact which resulted in the conclusion that for this case the RSB ansätze are required for precise approximation of the asymptotic performance.

Contributions and Organization

This paper determines the asymptotic distortion of the reconstruction scheme (2) when it is employed for recovering the source vector from the noisy sampling system (1) via the replica method. The distortion function, as well as the source distribution, is considered to be general, and the sampling matrix $\mathbf{A}$ belongs to a wide set of random ensembles. We deviate from the earlier replica analyses of compressive sensing systems by evaluating the general replica ansatz which includes all the possible structures for the replica correlation matrix. The generality of the replica ansatz enables us to determine the RS as well as RSB ansatz as special cases, and therefore, investigate the impact of symmetry breaking. The analytical results in special cases recover the earlier RS based studies of compressive sensing systems, e.g., [14], [16]–[18]. Moreover, our numerical investigations show that for the zero-norm reconstruction, the RS ansatz fails to predict the performance for relatively large compression rates while the RSB ansätze approximate the performance validly. An introduction to the replica method, is given through the asymptotic analyses in Section III.

Notation

We represent scalars, vectors and matrices with non-bold, bold lower case and bold upper case letters, respectively. A $k \times k$ identity matrix is shown by $\mathbf{I}_k$, and the $k \times k$ matrix with all entries equal to one is denoted by $\mathbf{1}_k$. $\mathbf{A}^{\mathsf{T}}$ indicates the Hermitian of the matrix $\mathbf{A}$. The set of real and integer numbers are denoted by $\mathbb{R}$ and $\mathbb{Z}$, and their corresponding non-negative subsets by superscript $+$. $\|\cdot\|$ and $\|\cdot\|_1$ denote the ℓ_2- and ℓ_1-norm respectively, and $\|\boldsymbol{x}\|_0$ represents the zero-norm defined as the number of nonzero entries. For a random variable x, p_x represents either the Probability Mass Function (PMF) or Probability Density Function (PDF), and F_x identifies the Cumulative Distribution Function (CDF). Moreover, E_x identifies mathematical expectation over x, and an expectation over all random variables involved in a given expression is denoted by E. For sake of compactness, the set of integers $\{1, \ldots, n\}$ is abbreviated as $[1:n]$ and a zero-mean and unit-variance Gaussian distribution is represented by $\phi(\cdot)$. Gaussian averages are shown as

$$\int (\cdot)\ \mathrm{D}z = \int (\cdot)\ \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\ \mathrm{d}z. \tag{4}$$

Whenever needed, we assume the support $\mathbb{X}$ to be discrete. The results, however, are in full generality and hold also for continuous distributions.

II. Problem Formulation

Suppose $\boldsymbol{y}_{k\times 1}$ is given by a sampling system as in (1) where

(a) $\boldsymbol{x}_{n\times 1}$ is an i.i.d. random vector with each entry being distributed with p_x over $\mathbb{X} \subseteq \mathbb{R}$.

(b) $\mathbf{A}$ is a $k \times n$ random matrix over $\mathbb{R}^{k\times n}$, such that its Gramian $\mathbf{J} := \mathbf{A}^{\mathsf{T}}\mathbf{A}$ has the eigendecomposition

$$\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{T}} \tag{5}$$

with $\mathbf{U}$ being an orthogonal Haar distributed matrix and $\mathbf{D}$ being a diagonal matrix of which the empirical CDF of eigenvalues (density of states) converges as $n \uparrow \infty$ to a deterministic CDF $\mathrm{F}_{\mathbf{J}}$.

(c) $\boldsymbol{z}_{k\times 1}$ is a real i.i.d. zero-mean Gaussian random vector with variance λ_0, i.e., $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \lambda_0\mathbf{I})$.

(d) The number of observations k is a deterministic function of the system dimension n such that

$$\lim_{n\uparrow\infty} \frac{k(n)}{n} = \frac{1}{\mathsf{r}} < \infty. \tag{6}$$

(e) $\boldsymbol{x}$, $\mathbf{A}$ and $\boldsymbol{z}$ are independent.

We reconstruct the source vector $\boldsymbol{x}$ from $\boldsymbol{y}$ as $\hat{\boldsymbol{x}} = \mathbf{g}(\boldsymbol{x})$ with $\mathbf{g}(\cdot)$ being defined in (2) and satisfies the following constraints.

(a) The penalty function $u(\cdot)$ decouples meaning that

$$u(\boldsymbol{v}) = \sum_{i=1}^{k} u(v_i). \tag{7}$$

(b) For a given vector $\boldsymbol{y}$, the objective function in (2) has a unique minimizer over the support $\mathbb{X}^n$.

Our goal is to determine the asymptotic average distortion D for this setup which is defined as the large limit of the expected average distortion defined in (3), i.e., $\mathsf{D} = \lim_{n\uparrow\infty} \mathsf{E}\, \mathsf{D}_n$. To do so, we represent D as the macroscopic parameter of a spin glass and invoke the replica method to determine that.

III. Statistical Mechanical Approach

Consider a spin glass with the Hamiltonian

$$\mathcal{E}(\boldsymbol{v}|\boldsymbol{y}, \mathbf{A}) = \frac{1}{2\lambda}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{v}\|^2 + u(\boldsymbol{v}) \tag{8}$$

for given $\boldsymbol{y}$ and $\mathbf{A}$. At the inverse temperature β, the microstate $\boldsymbol{v}$ is distributed with

$$\mathrm{p}^{\beta}(\boldsymbol{v}) = \frac{e^{-\beta\mathcal{E}(\boldsymbol{v}|\boldsymbol{y},\mathbf{A})}}{\sum_{\boldsymbol{v}} e^{-\beta\mathcal{E}(\boldsymbol{v}|\boldsymbol{y},\mathbf{A})}}. \tag{9}$$

By using a standard large deviation argument and defining

$$\mathsf{d}(\boldsymbol{v}; \boldsymbol{x}) = \sum_{j=1}^{n} \mathsf{d}(v_j; x_j), \tag{10}$$

it is shown that the asymptotic distortion reads

$$\mathsf{D} = \lim_{\beta\uparrow\infty} \lim_{n\uparrow\infty} \frac{1}{n}\, \mathsf{E}\left\{\mathsf{E}_{\boldsymbol{v}}^{\beta}\, \mathsf{d}(\boldsymbol{v}; \boldsymbol{x})\right\} \tag{11}$$

where $\mathsf{E}_{\boldsymbol{v}}^{\beta}$ takes expectation over the vector $\boldsymbol{v}$ with respect to (w.r.t.) p^{β}. (11) describes D as a macroscopic parameter of the spin glass specified by the Hamiltonian in (8). At this point, one utilizes a common trick in statistical mechanics which defines the "modified partition function" as

$$\mathcal{Z}(\beta, h|\boldsymbol{y}, \mathbf{A}) = \sum_{\boldsymbol{v}} e^{-\beta\mathcal{E}(\boldsymbol{v}|\boldsymbol{y},\mathbf{A}) + h\mathsf{d}(\boldsymbol{v},\boldsymbol{x})}, \tag{12}$$

and determines the macroscopic parameter as

$$\mathsf{D} = \lim_{\beta\uparrow\infty} \lim_{n\uparrow\infty} \frac{1}{n}\frac{\partial}{\partial h}\, \mathsf{E} \log \mathcal{Z}(\beta, h|\boldsymbol{y}, \mathbf{A})|_{h=0}. \tag{13}$$

(13) raises the nontrivial problem of determining a logarithmic expectation. Here, one may take a step further and employ the Riesz equality which for a given random variable t states

$$\mathsf{E} \log t = \lim_{m\downarrow 0} \frac{1}{m} \log \mathsf{E} t^m, \tag{14}$$

and write (13) as

$$\mathsf{D} = \lim_{\beta\uparrow\infty} \lim_{n\uparrow\infty} \lim_{h\downarrow 0} \lim_{m\downarrow 0} \frac{1}{n}\frac{\partial}{\partial h}\frac{1}{m} \log \mathsf{E}[\mathcal{Z}(\beta, h|\boldsymbol{y}, \mathbf{A})]^m. \tag{15}$$

Determining the right hand side (r.h.s.) of (15) faces two main difficulties. In fact, one needs to evaluate the moment for any real value of m (or at least in the right neighborhood of 0), and also take the limits in the order stated. This is where the replica method plays its role. It considers the expression under the logarithm in the r.h.s. of (15) as a function in terms of m, namely $f(m)$, and conjectures that

1) the analytic continuation of $f(\cdot)$ from $\mathbb{Z}^+$ onto $\mathbb{R}^+$ equals to $f(m)$ which intuitively states that the final expression of $f(m)$ determined for $m \in \mathbb{Z}^+$ is same as $f(m)$ for real values of m, and
2) the limits with respect to m and n exchange.

The conjecture is known as "replica continuity" and is where the replica method lacks rigorousness. By the replica continuity conjecture, $f(m)$ reads

$$\begin{aligned} f(m) &\coloneqq \mathsf{E}[\mathcal{Z}(\beta, h|\boldsymbol{y}, \mathbf{A})]^m \\ &= \mathsf{E} \prod_{a=1}^{m} \sum_{\boldsymbol{v}_a} e^{-\beta\mathcal{E}(\boldsymbol{v}_a|\boldsymbol{y},\mathbf{A}) + h\mathsf{d}(\boldsymbol{v}_a,\boldsymbol{x})}. \end{aligned} \tag{16}$$

IV. Main Results

Proposition 1 gives the general replica ansatz which only relies on the replica continuity conjecture. Before stating the proposition, let us define the R-transform.

Definition: Considering a random variable $t \sim \mathrm{p}_t$, the corresponding Stieltjes transform over the upper half complex plane is defined as $\mathrm{G}_t(s) = \mathsf{E}(t-s)^{-1}$. Denoting the inverse w.r.t. composition by $\mathrm{G}_t^{-1}(\cdot)$, the R-transform is given by

$$\mathrm{R}_t(\omega) = \mathrm{G}_t^{-1}(\omega) - \omega^{-1} \tag{17}$$

such that $\lim_{\omega\downarrow 0} \mathrm{R}_t(\omega) = \mathsf{E}t$. The definition can be also extended to matrix arguments. Assuming a matrix $\mathbf{M}_{n\times n}$ to have the eigendecomposition $\mathbf{M} = \mathbf{U}\, \mathrm{diag}[\lambda_1, \ldots, \lambda_n]\, \mathbf{U}^{\mathsf{T}}$, $\mathrm{R}_t(\mathbf{M})$ is then defined as $\mathrm{R}_t(\mathbf{M}) = \mathbf{U}\, \mathrm{diag}[\mathrm{R}_t(\lambda_1), \ldots, \mathrm{R}_t(\lambda_n)]\, \mathbf{U}^{\mathsf{T}}$.

Proposition 1 (General Replica Ansatz)**:** Let the linear system (1) fulfill the constraints of Section II. For non-negative integer m, define the function

$$\mathcal{D}(\beta, m) = \frac{1}{m}\, \mathsf{E}\mathsf{d}(\mathbf{v}; \mathbf{x}) \tag{18}$$

where $\mathbf{x}_{m\times 1}$ is a vector with all elements equal to the random variable $x \sim \mathrm{p}_x$, and $\mathbf{v}_{m\times 1} \in \mathbb{X}^m$ is a random vector with conditional distribution

$$\mathrm{p}_{\mathbf{v}|\mathbf{x}}^{\beta}(\mathbf{v}|\mathbf{x}) = \frac{e^{-\beta(\mathbf{x}-\mathbf{v})^{\mathsf{T}}\mathbf{T}\mathrm{R}_{\mathbf{J}}(-2\beta\mathbf{T}\mathbf{Q})(\mathbf{x}-\mathbf{v}) - \beta u(\mathbf{v})}}{\sum_{\mathbf{v}} e^{-\beta(\mathbf{x}-\mathbf{v})^{\mathsf{T}}\mathbf{T}\mathrm{R}_{\mathbf{J}}(-2\beta\mathbf{T}\mathbf{Q})(\mathbf{x}-\mathbf{v}) - \beta u(\mathbf{v})}}. \tag{19}$$

In (19), $\mathrm{R}_{\mathbf{J}}(\cdot)$ is the R-transform corresponding to $\mathrm{F}_{\mathbf{J}}$,

$$\mathbf{T}_{m\times m} = \frac{1}{2\lambda}\mathbf{I}_m - \beta\frac{\lambda_0}{2\lambda^2}\mathbf{1}_m, \tag{20}$$

and $\mathbf{Q}$ is the so-called replica correlation matrix which satisfies the fixed point equation

$$\mathbf{Q} = \mathsf{E}\ \mathsf{E}_{\mathrm{p}^{\beta}_{\mathbf{v}|\mathbf{x}}}(\mathbf{x}-\mathbf{v})(\mathbf{x}-\mathbf{v})^{\mathsf{T}} \tag{21}$$

where $\mathsf{E}_{\mathrm{p}^{\beta}_{\mathbf{v}|\mathbf{x}}}$ takes expectation over $\mathbf{v}$ w.r.t. $\mathrm{p}^{\beta}_{\mathbf{v}|\mathbf{x}}$. Then, under the replica continuity conjecture, the asymptotic average distortion is given by

$$\mathsf{D} = \lim_{\beta\uparrow\infty}\lim_{m\downarrow 0}\mathcal{D}(\beta, m). \tag{22}$$

Sketch of the proof: Starting from (16) and after evaluating the expectations w.r.t. $\boldsymbol{z}$ and $\mathbf{A}$, the left hand side (l.h.s.) of (16) is expressed in terms of the replica correlation matrix $\mathbf{Q}_{m\times m}$ whose entries are defined as

$$[\mathbf{Q}]_{ab} = \frac{1}{n}(\boldsymbol{x}-\boldsymbol{v}_a)^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{v}_b). \tag{23}$$

Taking the limits $n\uparrow\infty$ and $h\downarrow 0$, one is lead to use the saddle point method. Finally, using the law of large numbers, the equations in Proposition 1 are obtained. The detailed derivations are given in [25]. ■

Solving the fixed point equation (21) is notoriously difficult and possibly not of use, because it may depend in a non-analytic way on m. To address both issues, one restricts the search of the fixed point solutions to a small parameterized set of correlation matrices. In the sequel, we treat some of the well-known sets.

A. RS Ansatz

RS assumes that the valid solution of the fixed-point equation (21) is invariant under all permutations of the m replica indices, namely $\mathbf{\Pi}^{-1}\mathbf{Q}\mathbf{\Pi} = \mathbf{Q}$ for all permutation matrices $\mathbf{\Pi}$ taken from the symmetric group on $[1:m]$. This implies that $\mathbf{Q}$ is of the form

$$\mathbf{Q} = q\mathbf{1}_m + \frac{\chi}{\beta}\mathbf{I}_m \tag{24}$$

for some non-negative real q and χ. Indeed, this leads to an analytic expression for (18), and therefore, the limit in (22) is determined which concludes the following ansatz.

RS Ansatz: Define $\xi \coloneqq \lambda\ [\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})]^{-1}$ and f as

$$f \coloneqq \frac{1}{\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})}\sqrt{\frac{\partial}{\partial\chi}[(\lambda_0\chi-\lambda q)\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})]} \tag{25}$$

for some χ and q. Moreover, let

$$\mathrm{g}\,(x,z) \coloneqq \arg\min_{v}\left[\frac{1}{2\xi}|x+fz-v|^2+u(v)\right]. \tag{26}$$

Then, the RS prediction of D is given by

$$\mathsf{D} = \mathsf{E}\int \mathsf{d}\,(\mathrm{g}\,(x,z)\,;x)\ \mathrm{D}z, \tag{27}$$

for $x\sim\mathrm{p}_x$, and χ and q satisfying

$$\chi = \frac{\xi}{f}\ \mathsf{E}_x\int(\mathrm{g}\,(x,z)-x)\,z\ \mathrm{D}z,$$
$$q = \mathsf{E}_x\int(\mathrm{g}\,(x,z)-x)^2\ \mathrm{D}z. \tag{28}$$

The postulated symmetry of the ansatz, assumed in RS, does not necessarily hold, and therefore, the RS ansatz may fail to give a valid prediction of D.

B. RSB Ansätze

The RSB structures are constructed via Parisi's iterative scheme introduced in [21]. The scheme takes the RS correlation matrix as the starting point, and then recursively constructs new structures. After b steps of recursion, which are referred to as breaking steps, the correlation matrix reads

$$\mathbf{Q} = q\mathbf{1}_m + \sum_{\kappa=1}^{b} p_\kappa \mathbf{I}_{\frac{m\beta}{\mu_\kappa}}\otimes\mathbf{1}_{\frac{\mu_\kappa}{\beta}} + \frac{\chi}{\beta}\mathbf{I}_m \tag{29}$$

for some non-negative χ and q and sequences $\{\mu_\kappa\}$, and $\{p_\kappa\}$. Similar to the RS case, (29) leads to an analytic expression for $\mathcal{D}(\beta,m)$ which lets us determine D via Proposition 1. For sake of compactness, we first state the result for $b=1$, and then illustrate the generalized ansatz with b breaking steps.

One-step RSB Ansatz: Define $\xi \coloneqq \lambda\ [\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})]^{-1}$ and

$$f \coloneqq \frac{1}{\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})}\sqrt{\frac{\partial}{\partial\varrho}[\lambda_0\varrho+\lambda p-\lambda q]\mathrm{R}_{\mathbf{J}}(-\frac{\varrho}{\lambda})} \tag{30a}$$

$$w \coloneqq \frac{1}{\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})}\sqrt{\frac{\lambda}{\mu}\,\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})-\mathrm{R}_{\mathbf{J}}(-\frac{\varrho}{\lambda})]} \tag{30b}$$

for some real χ, q, μ and p, and $\varrho \coloneqq \chi+\mu p$; moreover let

$$\mathcal{K}(v,x,z,y) = \frac{1}{2\xi}\left[(x-v)^2+2\,(x-v)\,(fz+wy)\right]+u(v) \tag{31}$$

and define the functions

$$\mathcal{L}\,(x,z,y) = \min_{v}\mathcal{K}(v,x,z,y) \tag{32a}$$
$$\mathrm{g}\,(x,z,y) = \arg\min_{v}\mathcal{K}(v,x,z,y) \tag{32b}$$

Then, the asymptotic average distortion is given by

$$\mathsf{D} = \mathsf{E}\int \mathsf{d}\,(\mathrm{g}\,(x,z,y)\,;x)\,\mathcal{I}\,(x,z,y)\ \mathrm{D}y\mathrm{D}z$$

where $x\sim\mathrm{p}_x$ and $\mathcal{I}\,(x,z,y)$ reads

$$\mathcal{I}\,(x,z,y) = \frac{e^{-\mu\mathcal{L}(x,z,y)}}{\int e^{-\mu\mathcal{L}(x,z,y)}\mathrm{D}y}, \tag{33}$$

for χ, p and q which satisfy

$$\varrho = \frac{\xi}{f}\,\mathsf{E}\int(\mathrm{g}\,(x,z,y)-x)\,z\,\mathcal{I}\,(x,z,y)\,\mathrm{D}y\mathrm{D}z, \tag{34a}$$

$$\chi+\mu q = \frac{\xi}{w}\,\mathsf{E}\int(\mathrm{g}\,(x,z,y)-x)\,y\,\mathcal{I}\,(x,z,y)\,\mathrm{D}y\mathrm{D}z, \tag{34b}$$

$$q = \mathsf{E}\int \left(\mathrm{g}\left(x,z,y\right)-x\right)^2 \mathcal{I}\left(x,z,y\right)\mathrm{D}y\mathrm{D}z \tag{34c}$$

and μ which is a solution to

$$\frac{\mu p}{2\xi} = \frac{1}{2\lambda}\int_{\chi}^{\varrho} \mathrm{R}_{\mathbf{J}}(-\frac{\omega}{\lambda})\,\mathrm{d}\omega + \frac{\mu^2 w^2}{2\xi^2}(p-q) + \mathsf{E}\int \mathcal{I}\left(x,z,y\right)\log\mathcal{I}\left(x,z,y\right)\mathrm{D}y\mathrm{D}z. \tag{35}$$

The one-step RSB ansatz is further extended to more steps of breaking. In that case the asymptotic distortion is given by

$$\mathsf{D}=\mathsf{E}\int \mathsf{d}(\tilde{\mathrm{g}}(x,z,\{y_\kappa\}_1^b);x)\tilde{\mathcal{I}}(x,z,\{y_\kappa\}_1^b)\prod_{\kappa=1}^{b}\mathrm{D}y_\kappa\mathrm{D}z \tag{36}$$

where

$$\tilde{\mathcal{I}}(x,z,\{y_\kappa\}_1^b) = \prod_{\kappa=1}^{b}\left[\Lambda_\kappa(x,z,\{y_\varsigma\}_\kappa^b)\right]^{\mu_\kappa} \tag{37}$$

for a sequence of functions $\{\Lambda_\kappa(\cdot,\cdot,\{\cdot\})\}_{\kappa=1}^b$, and $\tilde{\mathrm{g}}(\cdot,\cdot,\{\cdot\}_1^b)$ being defined as in (32b) by replacing wy in (31) by $\sum_{\kappa=1}^b w_\kappa y_\kappa$ for some $\{w_\kappa\}_{\kappa=1}^b$. The explicit expression of the functions and scalar factors are omitted for sake of compactness, and are given in the extended version of the manuscript [25].

V. Least-Square Reconstruction Schemes

In compressive sensing, the source vector is supposed to be sparse which means that a certain fraction of entries are zero. To model the sparsity of the source, we set F_x to be

$$\mathrm{F}_x(x) = (1-s)\mathbf{1}\{x\geq 0\} + s\breve{\mathrm{F}}_x(x). \tag{38}$$

with $\mathbf{1}\{\cdot\}$ being the indicator function, for some CDF $\breve{\mathrm{F}}_x(x)$ and $0\leq s\leq 1$. By the law of large numbers, $\boldsymbol{x}$ in the asymptotic regime has $(1-s)n$ zeros and sn non-zero entries which are distributed w.r.t. CDF $\breve{\mathrm{F}}_x$. From the reconstruction point of view, several schemes can be considered by setting different forms of $u(\cdot)$ in (2). In this section, we consider the least-square based reconstruction schemes in compressive sensing and investigate the asymptotic performance. Throughout the investigations, we assume

- $\boldsymbol{x}$ is an i.i.d. zero-mean and unit-variance "sparse Gaussian" vector meaning that $\breve{\mathrm{F}}_x(x)$ in (38) is a zero-mean and unit-variance Gaussian CDF.
- the distortion function $\mathsf{d}(\cdot;\cdot)$ is of the form

$$\mathsf{d}(\hat{x};x) = (\hat{x}-x)^2 \tag{39}$$

  which determines the asymptotic average MSE.
- the sampling matrix is either an "i.i.d. random" or a "random projector" matrix. In the former case, the entries of $\mathbf{A}_{k\times n}$ are generated i.i.d. with zero-mean and variance k^{-1}. The asymptotic empirical eigenvalue CDF of the Gramian $\mathbf{J}$, in this case, follows the Marcenko-Pastur law, and therefore, the R-transform is given by

$$\mathrm{R}_{\mathbf{J}}(\omega) = \frac{1}{1-\mathsf{r}\omega}. \tag{40}$$

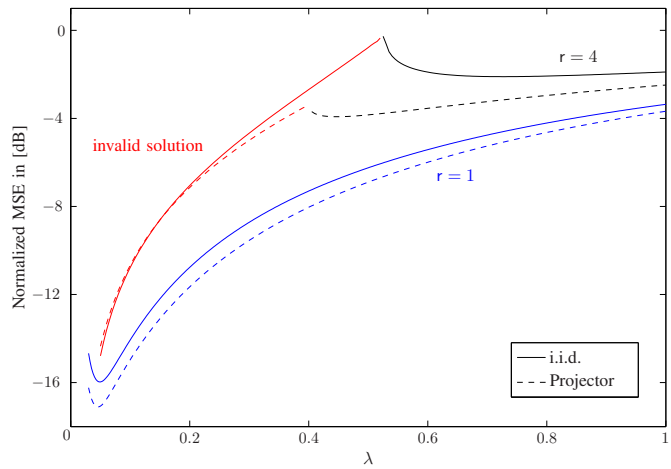

Fig. 1: RS-prediction of normalized MSE vs. the compression rate for the zero-norm reconstruction scheme. As the compression rate grows, the RS fixed-point equation gives invalid predictions. The sparsity factor is considered $s=0.1$ and the noise variance is set $\lambda_0=0.01$.

The latter case describes a sampling matrix in which the rows are orthogonal. In this case, the R-transform reads

$$\mathrm{R}_{\mathbf{J}}(\omega) = \frac{\omega-1+\sqrt{(\mathsf{r}\omega-1)^2+4\omega}}{2\omega}. \tag{41}$$

To study the asymptotics of least-square based reconstruction schemes, we need to set the penalty function to be one of ℓ_2-, ℓ_1-, or zero-norm functions.

(a) By setting the penalty function to be $u(v)=v^2/2$, the linear least-square scheme is recovered.
(b) For $u(v)=|v|$, the reconstruction scheme in (2) reduces to the LASSO [3] or basic pursuit denoising [26] scheme.
(c) By considering $u(v)=\mathbf{1}\{v\neq 0\}$ the zero-norm reconstruction scheme is obtained.

A. Numerical Results

Fig. 1 shows the RS prediction of the normalized MSE in terms of the tuning factor λ for the zero-norm reconstruction. The sparsity factor is considered to be $s=0.1$, and the noise variance is set to be $\lambda_0=0.01$. The curves have been sketched for both the i.i.d. random and projector sampling matrices at the compression rates $\mathsf{r}=1$ and $\mathsf{r}=4$. As the figure illustrates, RS fails to predict the normalized MSE at small values of λ for large compression rates. In fact, as the compression rate grows, the normalized MSE drops unexpectedly down for an interval of λ. This is due to the fact that the RS fixed point equations have either an invalid solution or no solution in this interval. In other words, the replica ansatz, for this regime of system parameters, does not exhibit symmetry, and therefore, the RS postulated structure for the replica correlation matrix does not lead to the true saddle-point. The result was earlier reported for the noiseless case in [24] where the authors showed that under a set of constraints the RS prediction is not valid for the zero-norm reconstruction.

To investigate the impact of RSB, the RS as well as one-step RSB prediction of the normalized MSE has been plotted

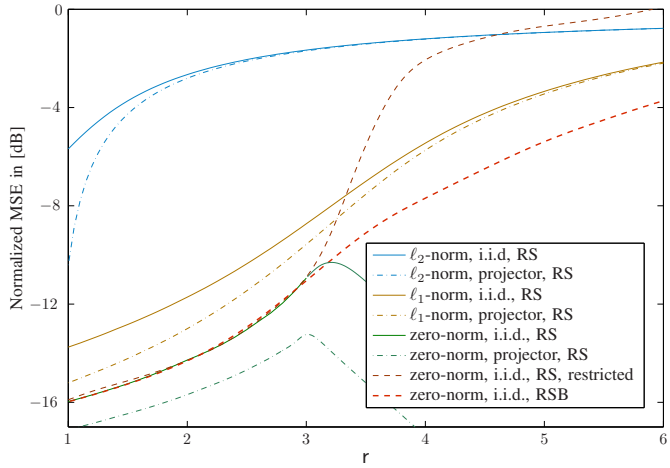

Fig. 2: RS and one-step RSB prediction of the normalized MSE vs. r for $s = 0.1$ and $\lambda_0 = 0.01$. At higher compression rates, the RS predicted MSE unexpectedly drops down. The one-step RSB, however, tracks the curve for ℓ_1-norm reconstruction within a gap.

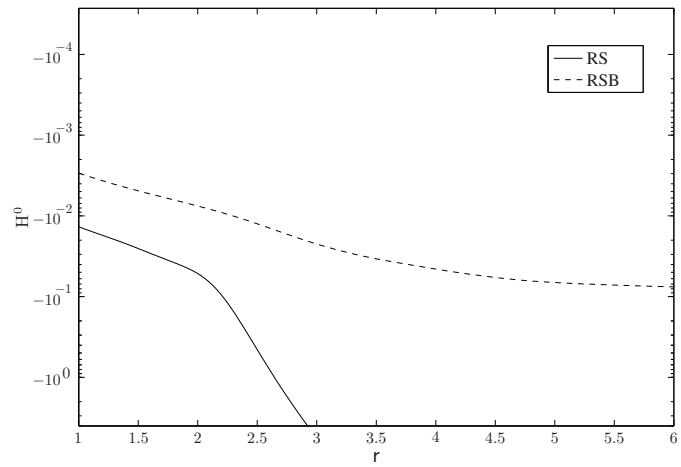

Fig. 3: RS and one-step RSB approximation of the zero-temperature entropy for the system setup corresponding to the i.i.d. RSB curve in Fig. 2. The figure confirms the better accuracy of the RSB prediction.

in terms of the compression rate in Fig. 2 for the zero-norm reconstruction when $s = 0.1$ and $\lambda_0 = 0.01$. The normalized MSE has been numerically minimized over λ. As a benchmark, the RS predicted curves for the ℓ_2- and ℓ_1-norm schemes have also been sketched. As the figure shows, the RS predicted MSE starts to decrease in higher compression rate regimes and even violates theoretical lower bounds. The one-step RSB ansatz, however, is consistent with theoretical bounds, and tracks the curve for the ℓ_1-norm scheme within a certain gap. For sake of comparison, we have also plotted the "restricted RS prediction". For this curve, we have minimized the RS-predicted normalized MSE within the interval of λ in which the RS ansatz is stable. As it is observed, the curve deviates the one-step RSB curve, as r grows large. It also violates the ℓ_1- and ℓ_2-norm curves at higher compression rates; the fact which indicates that the optimal tuning factor lies within the interval of λ with an unstable RS ansatz.

In order to study the accuracy of the one-step RSB further, we invoke the common consistency test based on the zero-temperature entropy H^0. In fact, considering the spin glass defined in (8), the distribution of the microstate at the zero temperature tends to an indicator function at the point which minimizes the Hamiltonian. Therefore, the entropy tends to zero. It has been, however, observed that in problems which RS clearly fails, the zero-temperature entropy is also predicted wrongly in the sense that it does not tend to zero, but becomes negative. It has been further shown that for these cases the zero-temperature entropy under the RSB ansätze takes values much closer to zero. Fig. 3 shows the zero-temperature entropy of the corresponding spin glass under both the RS and RSB assumptions versus the compression rate. The system setup has been considered as in Fig. 2 and H^0 has been determined at the tuning factors which minimize the one-step RSB predicted MSE. As the compression rate grows, the RS zero-temperature entropy drops down. The one-step RSB, however, gives a better approximation for H^0.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[4] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, 1975.

[5] M. Mézard and G. Parisi, "A replica analysis of the travelling salesman problem," *Journal de Physique*, vol. 47, no. 8, pp. 1285–1296, 1986.

[6] Y. Fu and P. W. Anderson, "Application of statistical mechanics to NP-complete problems in combinatorial optimisation," *Journal of Physics A: Mathematical and General*, vol. 19, no. 9, p. 1605, 1986.

[7] H. Nishimori, *Statistical physics of spin glasses and information processing: an introduction*. Clarendon Press, 2001, no. 111.

[8] A. Montanari, "Turbo codes: The phase transition," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 18, no. 1, pp. 121–136, 2000.

[9] T. L. Watkin, A. Rau, and M. Biehl, "The statistical mechanics of learning a rule," *Reviews of Modern Physics*, vol. 65, no. 2, p. 499, 1993.

[10] Y. Kabashima and D. Saad, "Statistical mechanics of error-correcting codes," *EPL (Europhysics Letters)*, vol. 45, no. 1, p. 97, 1999.

[11] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Transactions on Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, 2002.

[12] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Transactions on Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.

[13] R. R. Müller, "On channel capacity, uncoded error probability, ML-detection, and spin glasses," in *Proc. Workshop on Concepts in Information Theory*. Citeseer, pp. 79–81, 2002.

[14] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *47th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, pp. 52–59, 2009.

[15] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Transactions on Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, 2012.

[16] S. Rangan, A. K. Fletcher, and V. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," in *IEEE Transactions on Inf. Theory*, 2012, pp. 1902–1923.

[17] M. Vehkaperä, Y. Kabashima, and S. Chatterjee, "Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing," *IEEE Transactions on Inf. Theory*, vol. 62, no. 4, pp. 2100–2124, 2016.

[18] A. M. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Transactions on Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, 2013.

[19] B. M. Zaidel, R. Müller, A. L. Moustakas, and R. de Miguel, "Vector precoding for gaussian MIMO broadcast channels: Impact of replica symmetry breaking," *IEEE Transactions on Inf. Theory*, vol. 58, no. 3, pp. 1413–1440, 2012.

[20] R. Müller, D. Guo, and A. L. Moustakas, "Vector precoding for wireless MIMO systems and its replica analysis," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 530–540, 2008.

[21] G. Parisi, "A sequence of approximated solutions to the SK model for spin glasses," *Journal of Physics A: Mathematical and General*, vol. 13, no. 4, p. L115, 1980.

[22] A. Bereyhi, R. Müller, and H. Schulz-Baldes, "RSB decoupling property of MAP estimators," in *IEEE Information Theory Workshop (ITW)*. pp. 379–383, 2016.

[23] M. A. Sedaghat, A. Bereyhi, and R. Mueller, "LSE precoders for massive MIMO with hardware constraints: Fundamental limits," *arXiv preprint arXiv:1612.07902*, 2016.

[24] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on ℓ_p-norm minimization," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 09, p. L09003, 2009.

[25] A. Bereyhi, R. R. Müller, and H. Schulz-Baldes, "Statistical mechanics of MAP estimation: General replica ansatz," *arXiv preprint arXiv:1612.01980*, 2016.

[26] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.